

D2

(12) UK Patent Application (19) GB (11) 2 354 388 (13) A

(43) Date of A Publication 21.03.2001

(21) Application No 9916337.0

(22) Date of Filing 12.07.1999

(71) Applicant(s)

Independent Television Commission
(Incorporated in the United Kingdom)
33 Foley Street, LONDON, W1P 7LB, United Kingdom

(72) Inventor(s)

David Harrison
Nicolas Kim Lodge

(74) Agent and/or Address for Service

Boult Wade Tennant
Verulam Gardens, 70 Gray's Inn Road, London,
WC1X 8BT, United Kingdom

(51) INT CL⁷

H04N 5/262

(52) UK CL (Edition S)

H4F FGG F27V F70

(56) Documents Cited

GB 2050748 A GB 2014015 A EP 0276800 A2

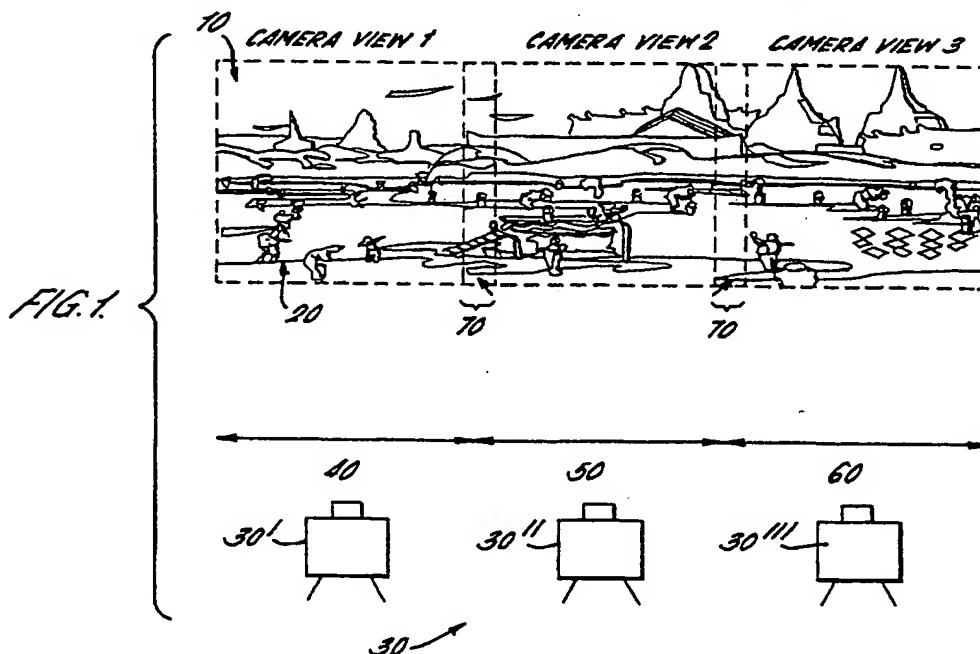
(58) Field of Search

UK CL (Edition S) H4F FAAE FAAX FGG FGH FGS FGT
INT CL⁷ H04N 5/262
ONLINE: WPI EPODOC JAPIO

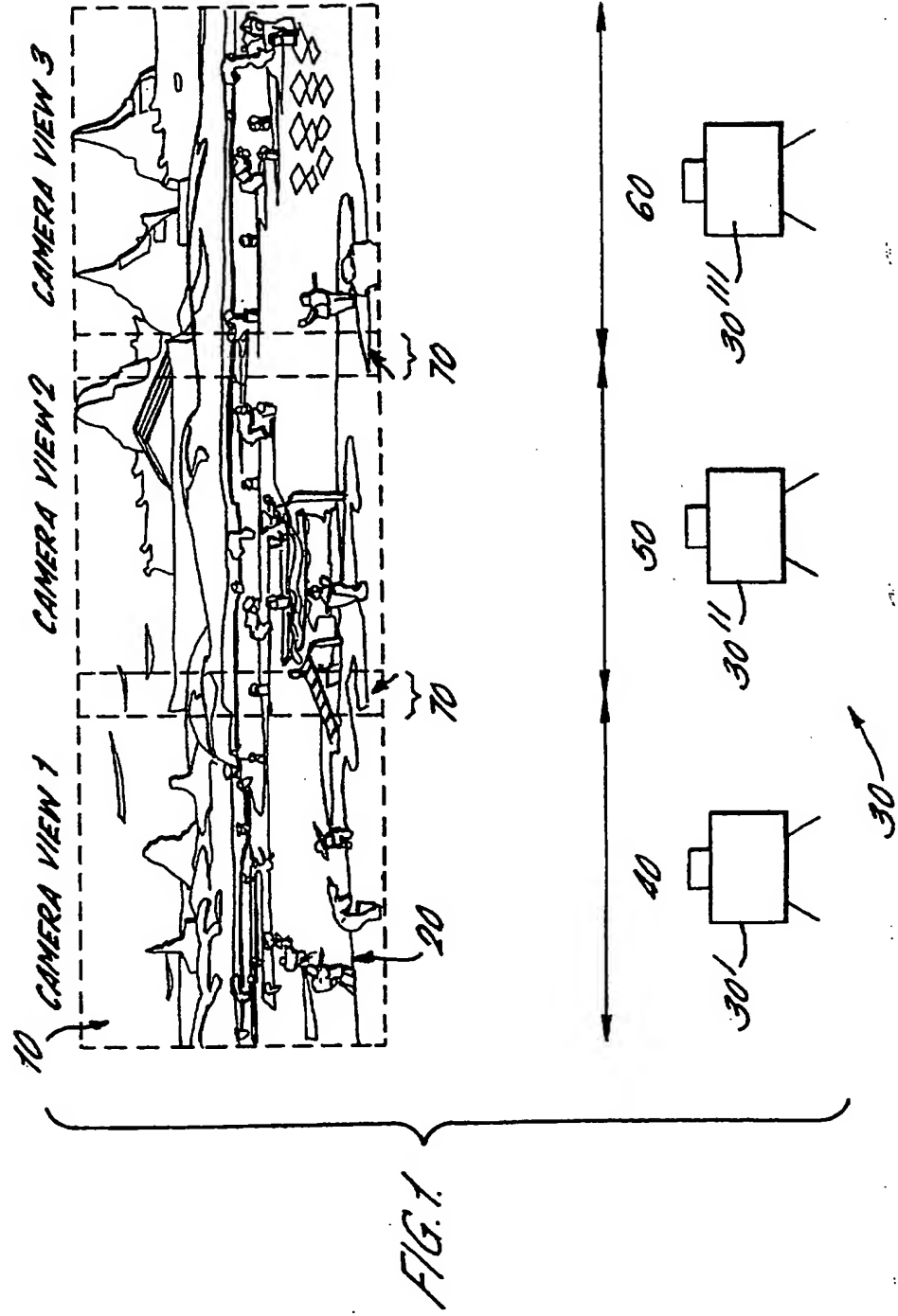
(54) Abstract Title

System and method for capture, broadcast and display of moving images

(57) A number of moving sub-images are captured of a large, or panoramic, image using more than one camera. Each camera points to an adjacent part of the large image. The sub-images are separately encoded, and can be multiplexed and compressed before transmission. Upon receiving the signal, at least a part of the panoramic image is displayed following decoding each signal representative of a sub-image. The signals are relayed to adjacent parts of a large display, optionally adjacent video or projection screens. In another embodiment, especially suitable for concerts or sports events, only a part of the panoramic moving image relating to the action is broadcast in real time. Other parts of the image relating to the crowd and/or sky is displayed as a prerecorded animated loop. This can allow the use of a single camera to capture the action.

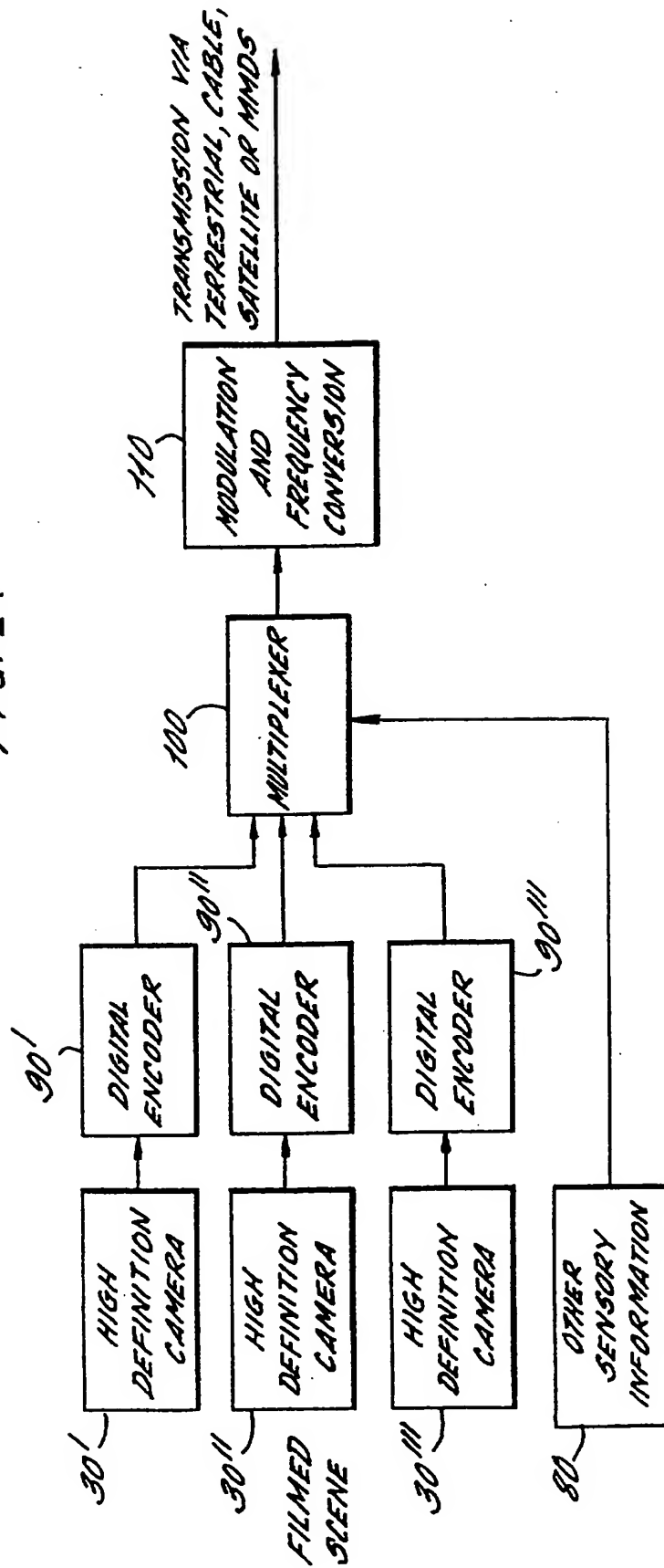


GB 2 354 388 A



000000

FIG. 2.



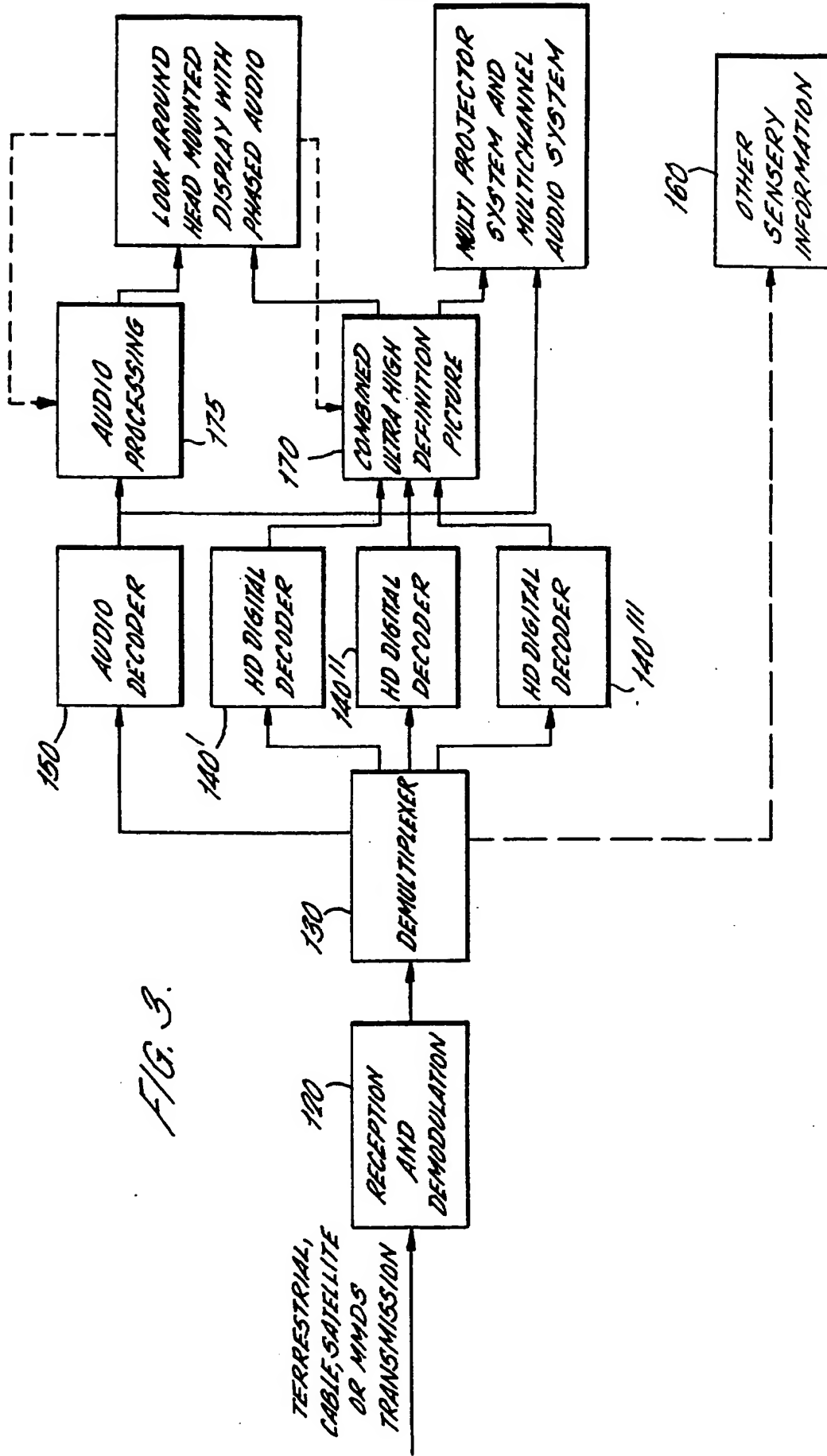


FIG. 3.

FIG. 4.

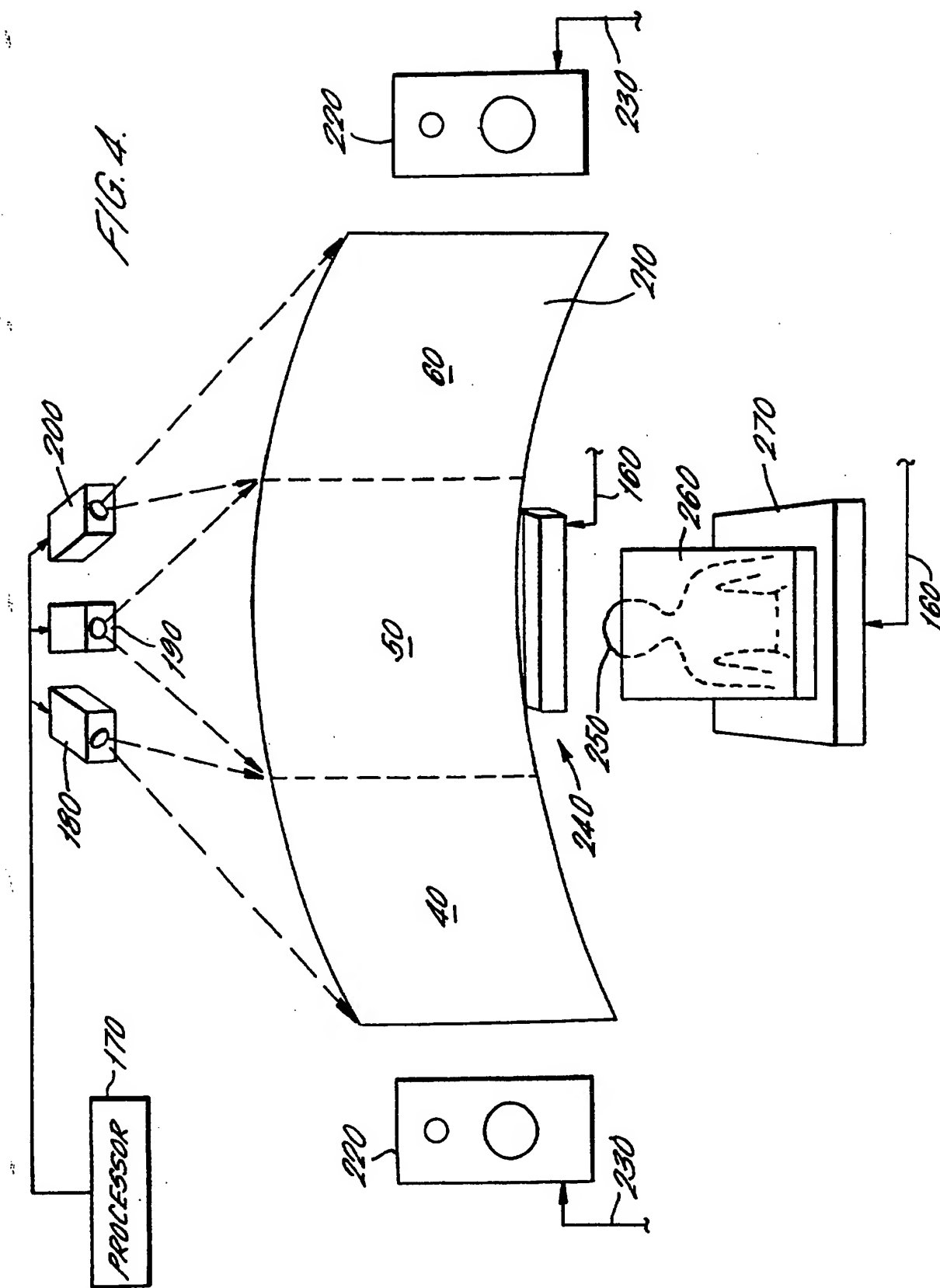


FIG. 5A.

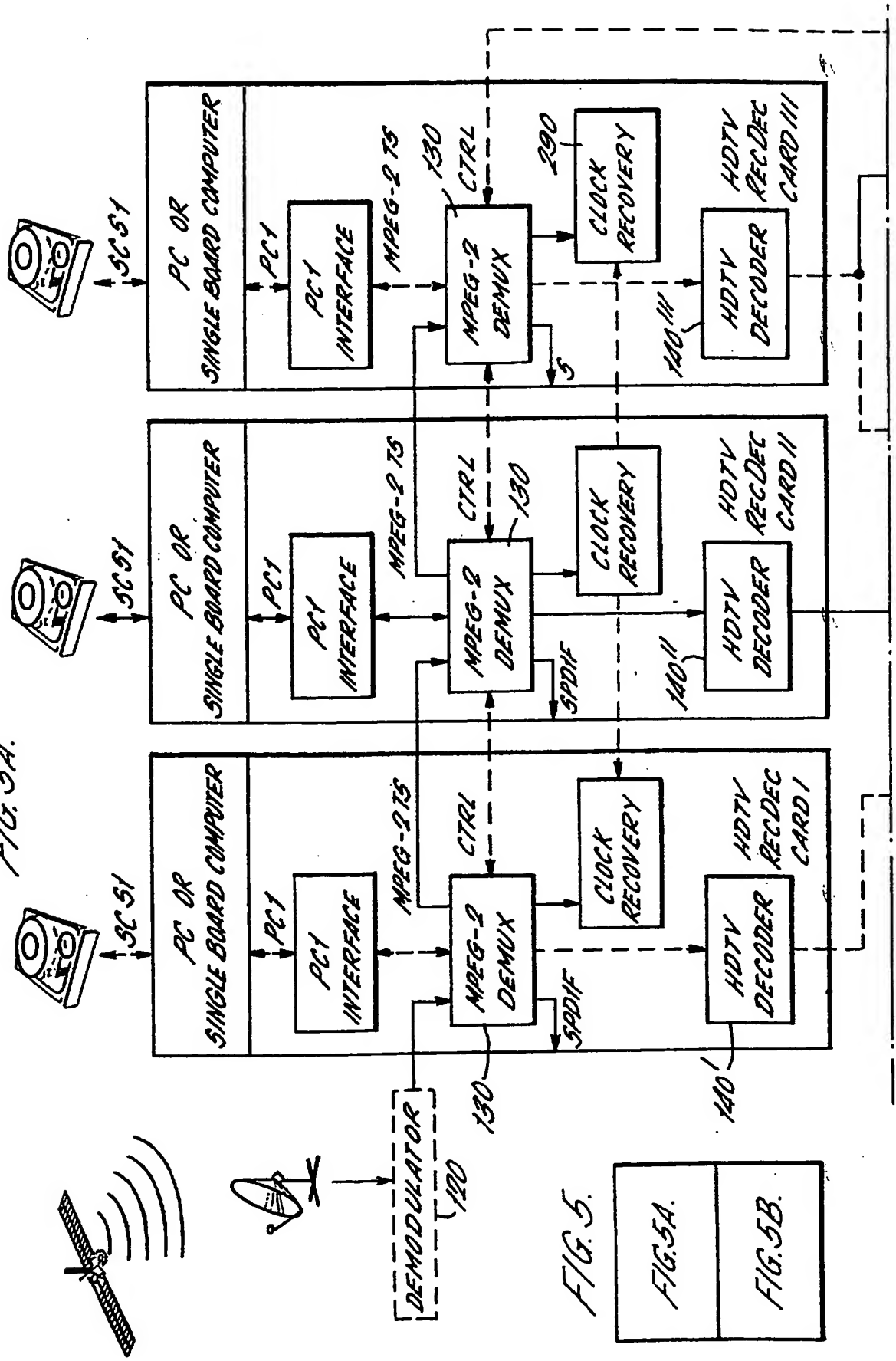
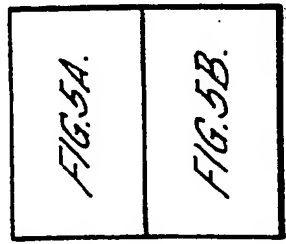


FIG. 5.



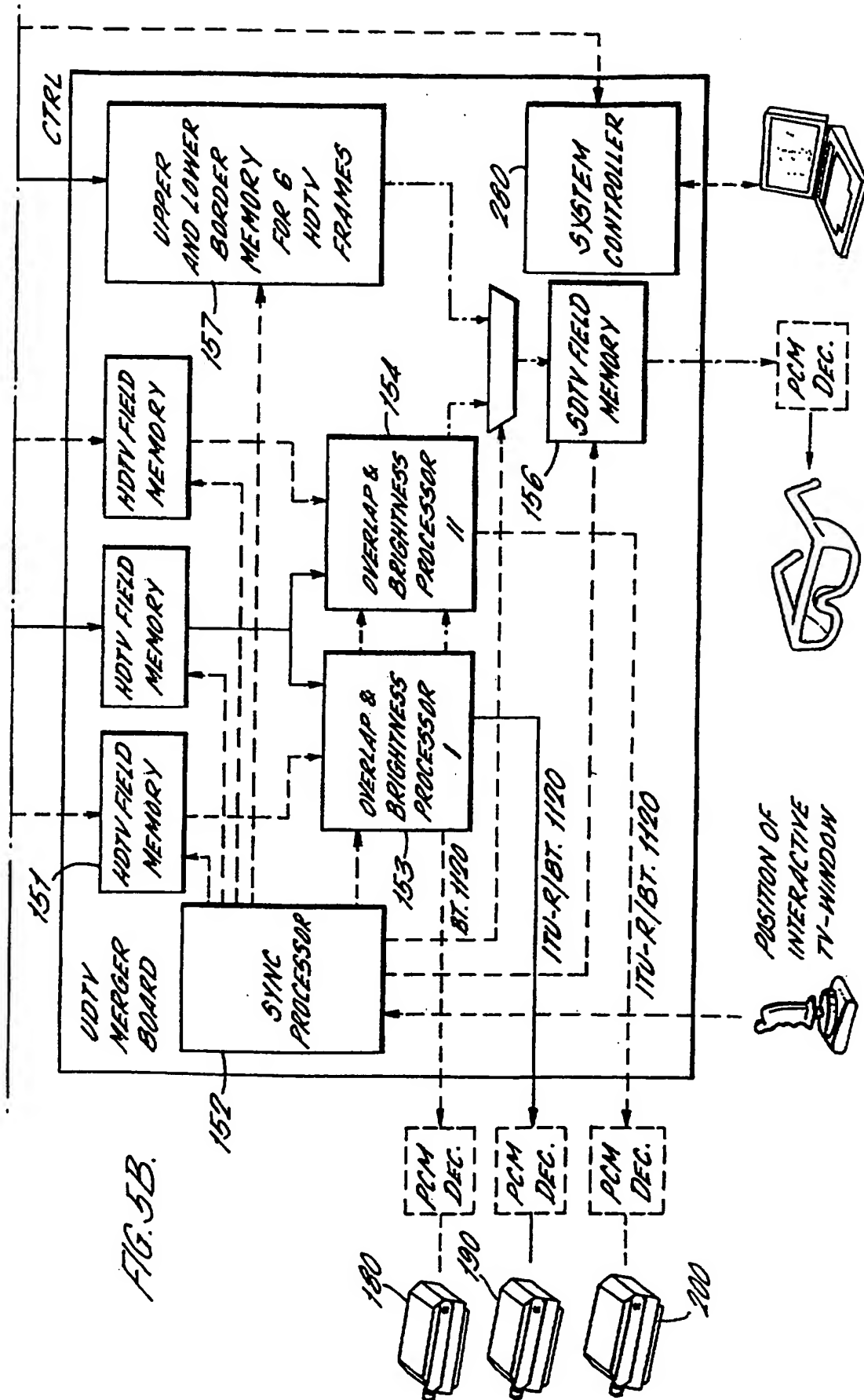


FIG. 5B.

FIG. 6.

300

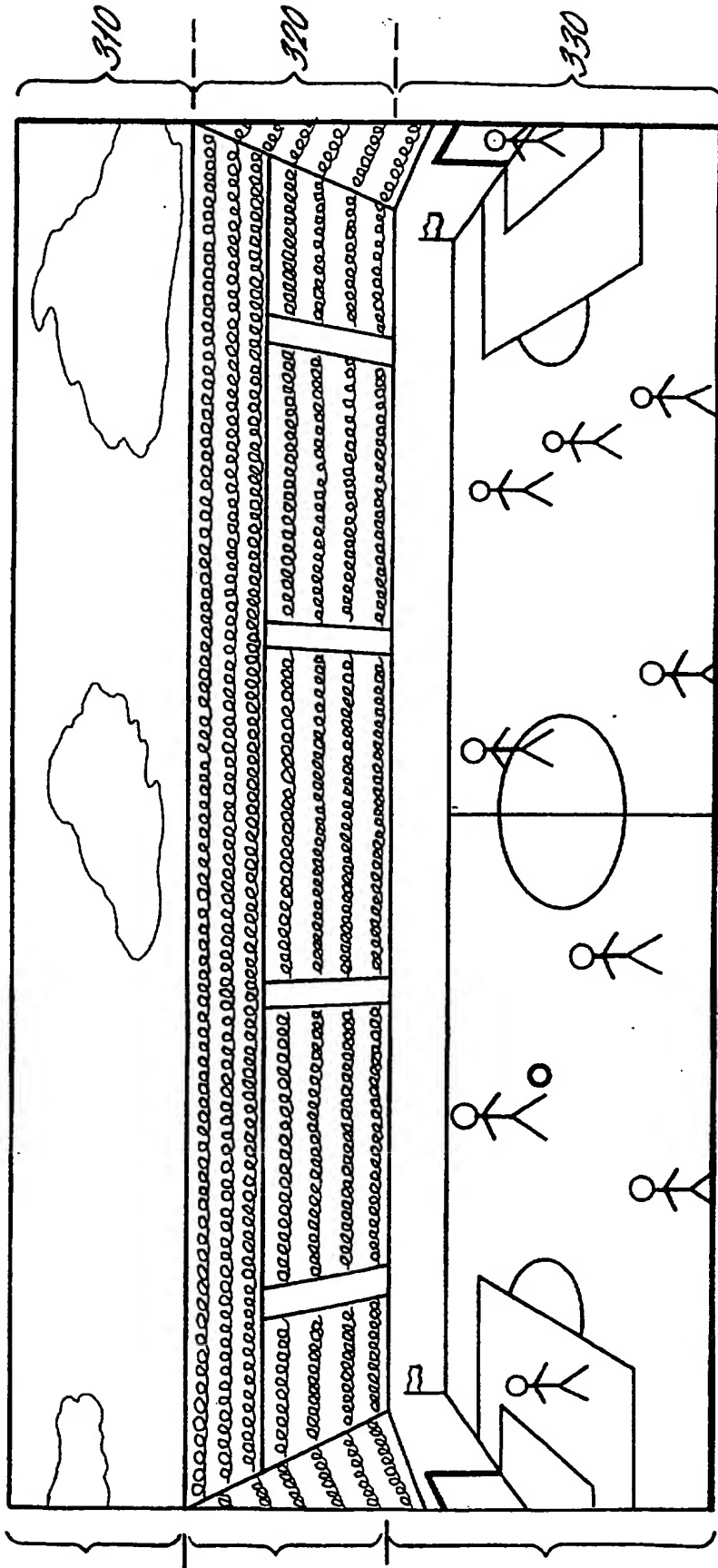


FIG. 7.

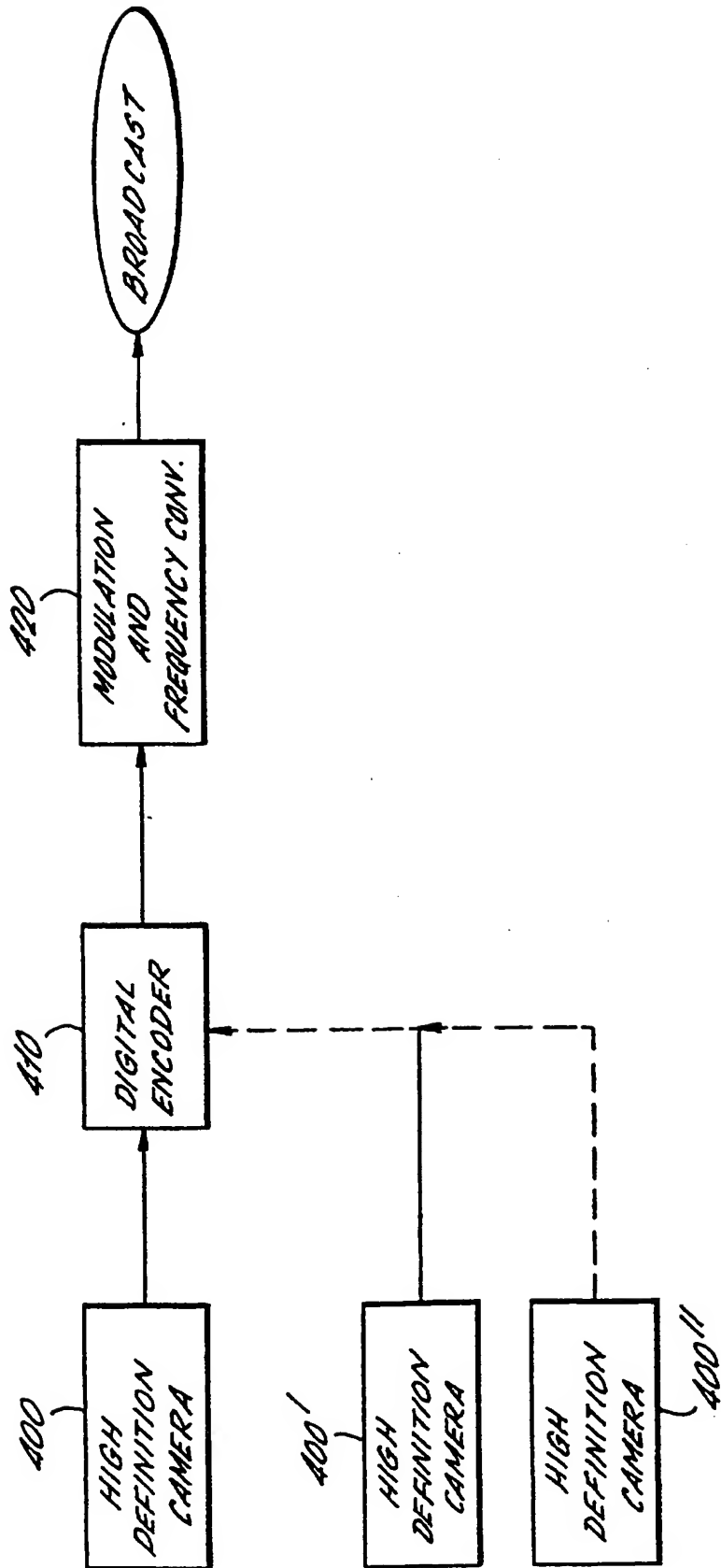
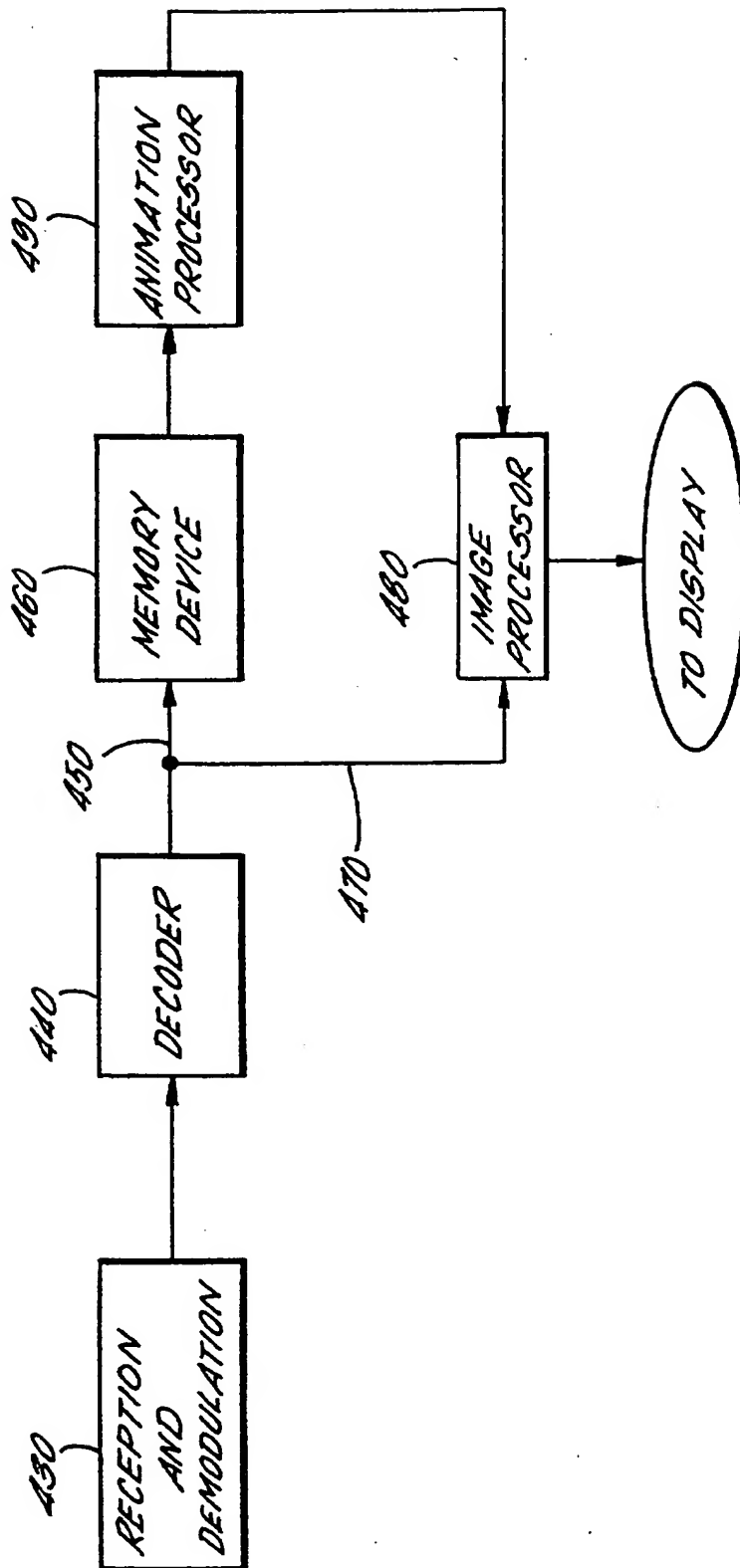


FIG. 8.



System and Method for Capture,
Broadcast and Display of Moving Images

5 This invention relates to a system and a method
for the capture, broadcast and display of moving
images such as audiovisual television images.

Audiovisual entertainment systems have existed
for many years; television and cinema, for example,
have been commercially available for nearly a century.
10 Over that period, various improvements have been
brought in, such as colour images and, latterly, Nicam
multichannel stereo which has improved the spatial
distribution of sound relative to a stationary viewer.
This in turn has lent a more accurate impression of
15 motion to accompanying images.

Whilst such forms of media provide a degree of
enjoyment, in their basic form they cannot be
considered particularly lifelike. Several attempts
have been made over the years to address this
20 limitation. In cinematography, for example,
stereoscopic or '3-D' movies have been displayed.
Here, two cameras are locked together but spatially
displaced relative to the field of view. By filtering
the red part of the visible spectrum on one of the
25 cameras, and the green part on the other complementary
camera, and then displaying the two images together,
an impression of depth of view of parallax is given.
However, to obtain this, it is necessary that the
viewer should wear a red filter over one of his or her
30 eyes and a green filter over the other eye. The costs
associated with filming two virtually identical images
and providing wearers with special colour filter
spectacles has meant that the system has never been

particularly popular.

Another cinematographic technique developed to improve the sense of realism is to capture a moving image from the perspective of a participant, over a 180° field of view; for example, a camera may be mounted upon a roller coaster in a position where a rider of the roller coaster might sit, the camera having a very wide angle lens. The resulting image is displayed on a large hemispherical screen under which the viewers are seated, and a sense of "immersion" within the image is achieved. The drawback with this technique is the cost and space necessary to provide a suitably large hemispherical screen.

The recent dramatic increase in computer processing power has also led to the development of so-called virtual reality devices. In one such device, the user wears a headset. Sounds and computer-generated images are fed into the headset so that the user has the sensation of being 'surrounded' by the resulting audio and video images. Sensors may be fitted to the headset so that movement of the head causes the images and apparent source of sounds to move in consequence. Although enjoyable, the computer-generated images are still clearly perceivable as such and cannot be considered particularly lifelike.

There have been few attempts at a more realistic or "immersive" approach in television. Apart from the drawbacks already set out, a primary problem with television is that images and sounds must be broadcast (i.e., transmitted and received) which immediately puts a finite limit upon the amount of information that can be sent. Even with the advent of compressed

digital signals, and cable transmission, in order to generate the type of images and sounds necessary for a realistic, "immersive" television broadcast using existing techniques would require an impractically large bit rate capacity.

It is an object of the present invention to provide a system and method for capture, transmission and display of moving images which alleviates these problems with the prior art.

In its broadest sense, the invention provides a method of capturing, broadcasting and displaying an image, comprising: (a) capturing, with n separate cameras, where n is a positive integer greater than one, n respective frames of subimages obtained from separate spatially adjacent subdivisions of the image; (b) digitally encoding said n captured arrays to generate a corresponding number, n , of separate encoded frames of subimages; (c) compressing the n encoded frames of subimages; (d) multiplexing the n encoded frames of subimages to form a composite encoded image signal; (e) transmitting the composite encoded image signal; (f) receiving the composite encoded image signal; (g) demultiplexing the composite encoded image signal to reconstitute the said number, n , of separate encoded frames of subimages; and (h) decoding and decompressing the n encoded frames of subimages such that the resultant frames of subimages may be displayed spatially adjacent to one another to provide a facsimile of the captured image.

A system for capturing, broadcasting and displaying an image according to the invention comprises a plurality, n , of separate cameras, where n is a positive integer greater than 1, arranged

respectively to capture n separate frames of subimages obtained from separate spatially adjacent subdivisions of the image; an encoder for digitally encoding said n captured arrays to generate a corresponding number, n , of separate encoded frames of subimages; a processor for compressing the n encoded frames of subimages; a multiplexer for multiplexing the n encoded frames of subimages to form a composite encoded image signal; a transmitter for transmitting the composite encoded image signal; a receiver for receiving the composite encoded image signal; a demultiplexer for demultiplexing the composite encoded image signal to reconstitute the said number, n , of separate encoded frames of subimages; a decoder for decoding and decompressing the n encoded frames of subimages; and a display for displaying at least some of the n decoded frames of subimages spatially adjacent to one another to provide a facsimile of at least part of the captured image.

More specifically, according to a first aspect of the present invention, there is provided a method of capturing, broadcasting and displaying a moving image, comprising: (a) capturing, with n separate cameras, where n is a positive integer greater than 1, n respective separate arrays of temporally successive still image subframes, each of the n arrays being captured from separate spatially adjacent subdivisions of the moving image respectively; (b) digitally encoding said n captured arrays to generate a corresponding number, n , of separate encoded arrays, each of which contains subframe image data; (c) compressing the size of at least some of the n digitally encoded arrays; (d) multiplexing the n

encoded arrays to form a composite encoded image
signal; (e) transmitting the composite encoded image
signal; (f) receiving the composite encoded image
signal; (g) demultiplexing the composite encoded image
5 signal to reconstitute the said number, n , of encoded
arrays; (h) decoding and decompressing the n encoded,
compressed arrays to reconstitute the n captured
arrays obtained from the original moving image; and
(i) displaying the n captured arrays in spatial order
10 such that a facsimile of at least a spatial part of
the original moving image may thereby be represented;
the subframe image data of the encoded arrays being
such that, over a given time period, the sum of the
compressed sizes of the arrays is the same as or less
15 than the size of the data contained within the largest
of the n uncompressed arrays over that time period.

The method of the invention thus allows a high
definition wide screen image to be broadcast using the
same techniques (cable, satellite or digital
20 terrestrial, for example) as are presently available,
but importantly without requiring additional bit rate
capacity. The wide screen image is in fact the
juxtaposition of n subimages, each of which may be
obtained for example using HDTV cameras, and projected
25 again using HDTV projectors.

The ability to compress the arrays of temporally
successive still image subframes depends upon the
content of the arrays being highly compressible. There
are several ways of ensuring that the criterion that,
30 over a given time period, the sum of the compressed
sizes of the arrays is at least as small as the size
of the data contained within the largest of the n
uncompressed encoded arrays over that time period, is

met.

For example, in a preferred embodiment, there are three arrays, captured with three separate cameras. If one assumes that the amount of data in each encoded array is roughly the same, then in order to broadcast over standard bit rate capacity channels, the compression of each array over a given time needs to be, on average, down to one third of the size of each of the three encoded arrays before compression (since each array is then effectively "the largest"). This does not necessarily mean that all three encoded arrays in this example must be compressed to one third or less of their original sizes. One of the images may be compressed more than the other two, provided that the above criterion is still met.

In one preferred embodiment, the compression may be accomplished using MPEG-2. This permits significant compression when the data in temporally successive image subframes in a given array is substantially the same.

Most of the coding requirements of MPEG-2 derive from the motion vectors (that is, the vectors which describe the movement between one frame and a subsequent or previous frame). When the data in temporally successive image subframes in a given array is substantially the same (that is, little motion vector encoding is necessary), noise reduction algorithms may be applied to the data with some accuracy. For example, it may generally be assumed that the data difference between frames should be relatively minimal. A low noise threshold may then be set, and any apparent differences between frames which exceed this threshold is assumed to have arisen from

noise rather than actual motion. Of course, the algorithm is in reality adaptive so that, should significant changes in the coded captured scene actually occur, the algorithm will not dismiss these changes as only noise.

The method is particularly suited to the capture, broadcast and display of live events such as football matches or concerts. By employing an array of fixed view, fixed focus cameras, filming across the pitch or towards the stage respectively, a viewer of the ultimate image is then given the impression of being present at the event. The bulk of the image, in fact, does not change with time in such circumstances. For example, the pitch or stage, and the surrounding crowds, lights, stadium and sky will (broadly) not change or move over the course of a game or performance. It is only the people on the pitch or stage, representing a small fraction of the overall image, that will change. Thus, a high degree of image compression is possible.

Preferably, the method further comprises capturing an audio soundtrack in synchronism with the n arrays of temporally successive image subframes; digitally encoding the said audio soundtrack; multiplexing the encoded audio soundtrack with the n encoded arrays such that the resultant composite encoded image signal includes both audio and video data; demultiplexing the composite encoded image signal after the said steps (e) and (f) of transmitting and receiving, to reconstitute both the n encoded arrays and the encoded audio soundtrack; decoding the said audio soundtrack; and playing back the said decoded audio soundtrack in synchronism with

the n arrays of moving images. Time codes may be employed to ensure correct synchronisation with the video signals when reconstructed and displayed.

5 The invention also extends, in a further aspect, to a system for capturing, broadcasting and displaying a moving image, comprising: a plurality, n , of separate cameras, where n is a positive integer greater than 1, arranged respectively to capture n separate arrays of temporally successive still image subframes, each of the arrays being captured from
10 separate spatially adjacent subdivisions of the moving image respectively; an encoder arranged to digitally encode the n captured arrays, to generate a corresponding number, n , of separate encoded arrays, each of which contains subframe image data; a
15 processor, arranged to compress the size of at least some of the n digitally encoded arrays; a multiplexer, arranged to multiplex the n encoded arrays to form a composite encoded image signal; a transmitter arranged
20 to transmit the composite encoded image signal; a receiver arranged to receive the composite encoded image signal; a demultiplexer for demultiplexing the composite encoded signal to reconstitute the said number, n , of encoded arrays; a decoder arranged to
25 decode and decompress the n encoded, compressed arrays to reconstitute the n captured arrays obtained from the original moving image; and a display for displaying the n captured arrays in spatial order such that a facsimile of at least a spatial part of the
30 original moving image may thereby be represented; the subframe image data of the encoded arrays being such that, over a given time period, the sum of the compressed sizes of the arrays is the same as, or less

than, the size of the data contained within the largest of the n uncompressed encoded arrays over that time period.

5 In a preferred embodiment, the n cameras are HDTV cameras. Thus, n high definition subframes can be obtained and compressed and broadcast. The ultimate displayed image, when reconstructed, appears as a single, wide view, high definition moving image.

10 The system may further comprise a microphone for capturing an audio soundtrack in synchronism with the n arrays; a sound encoder for digitally encoding the said audio soundtrack; and a sound reproduction device, wherein the multiplexer is further arranged to multiplex the encoded audio soundtrack with the n encoded, compressed arrays such that the resultant composite encoded image signal includes both audio and video data, the demultiplexer is further arranged to demultiplex the composite encoded image signal following reception thereof by the receiver, to
20 reconstitute both the n encoded arrays and the encoded audio soundtrack, and the decoder is further arranged to decode the said audio soundtrack such that the sound reproduction means may play back the said decoded audio soundtrack in synchronism with the n
25 arrays.

In one embodiment, the display may include a plurality, n , of projectors and a screen, each of the n projectors being arranged to project a respective one of the n arrays onto a part of the screen, the n
30 arrays being projected in the spatial order in which they were captured, such that the facsimile of all of the original moving image is represented upon the screen. In that case, the n projectors may be arranged

to project an array in HDTV format onto the screen.
In the alternative, one projector may project all of
the n arrays, again in spatial order, onto the screen.

Alternatively, the display may include a head
5 mounted viewing device arranged to display a spatially
limited portion of the original moving image, movement
of the head-mounted display device causing different
spatially limited parts of the original moving image
to be displayed.

10 According to a further aspect of the present
invention, there is provided a method of capturing,
broadcasting, and displaying a moving image,
comprising: capturing, with a camera, a substantially
temporally continuous stream of first still image
15 subframes from a first spatial part of the moving
image; the stream of first still image subframes
representing a substantially temporally continuous
motion picture image; digitally encoding the captured
first still image subframes; capturing, with a camera,
20 a finite series of second still image subframes from a
second spatial part of the moving image, the finite
period of time being short in relation to the time
over which the substantially continuous stream of
second still image subframes is captured, the finite
25 series of second still image subframes representing a
finite length motion picture image; digitally encoding
the said finite series of second still image
subframes; transmitting the said first and second
encoded still image subframes; receiving the said
30 first and second encoded still image subframes;
decoding the first and second encoded still image
subframes; providing a display; displaying the said
first still image subframes within a first spatial

part of the display such that the substantially temporally continuous motion picture image is represented; and repeatedly displaying the said second still image subframes within a second spatial part of the display such that the finite length motion picture image is represented as an animated loop; the display thereby displaying a facsimile of at least a spatial part of the original moving image.

This approach also addresses the issue of limited bit rate capacity transmission and allows a large field of view to be displayed whilst at the same time requiring only standard bit rate capacity transmission.

The method again finds particular application for live events such as concerts or football matches.

The invention also extends to a system for capturing, broadcasting and displaying a moving image, comprising: a first camera, arranged to capture a substantially temporally continuous stream of first still image subframes from a first spatial part of the moving image, the stream of first still image subframes representing a substantially temporally continuous motion picture image; a second camera, arranged to capture a finite series of second still image subframes from a second spatial part of the moving image, the finite period of time being short in relation to the time over which the substantially continuous stream of second still image subframes is captured, the finite series of second still image subframes representing a finite length motion picture image; an encoder arranged to digitally encode the captured stream of first still image subframes and to digitally encode the captured finite series of second

still image subframes; a transmitter for transmitting the said first and second encoded still image subframes; a receiver for receiving the said first and second encoded still image subframes; a decoder
5 arranged to decode the first and second encoded still image subframes; a display having first and second spatial parts, the system being arranged to display the said first still image subframes within the first spatial part of the display such that the
10 substantially temporally continuous motion picture image is represented and to repeatedly display the said second still image subframes within the second spatial part of the display such that the finite length motion picture image is represented as an
15 animated loop; the display thereby displaying a facsimile of at least a part of the original moving image.

Preferably, the encoder receives and encodes the finite series of second still image subframes prior to
20 receipt and encoding of the stream of first still image subframes, and in which the transmitter transmits the second still image subframes and the receiver receives the second still image subframes before transmitting and receiving the first still
25 image subframes respectively, the system further comprising: a storage device local to the receiver, the storage device being arranged to receive and store the said finite series of second still image subframes; and a processor arranged to receive the
30 finite series of second still image subframes from the storage device and to animate the said finite series as an animated loop in synchronism with the temporally continuous motion picture subsequently received by the

receiver. In other words, part of the displayed image is actually animated locally which prevents all of the information from being broadcast simultaneously. This further reduces the bit rate capacity requirements of the broadcast.

For example, one camera may first capture images of, say, a crowd, over a short (few seconds) period of time. This can be encoded and broadcast initially, and then recorded locally to the receiver for local animation. Once the short loop to be animated has been broadcast, the same camera may then switch over to capture "live" action such as of a game of football for live broadcast. The display can receive both the live action, in real time, as well as the animated images of the crowd and even the sky juxtaposed therewith.

It will be appreciated that the various aspects of the invention are not mutually exclusive. Therefore, the use of segmentation to generate live and animated images after broadcast can be employed in combination with the use of n separate cameras, each obtaining separate images to be compressed and broadcast as a multiplexed signal before reconstruction into a composite image at the display.

The invention may be put into practice in a number of ways, some of which will now be described by way of example only and with reference to the drawings in which:

Figure 1 shows a schematic diagram of a first typical wide angle scene to be captured, broadcast and displayed;

Figure 2 shows a block diagram of a first part of the system according to a first embodiment of the

invention for capture and transmission of images of the parts of the scene of Figure 1.

Figure 3 shows a block diagram of a second part of the system according to the first embodiment of the invention for reception and display of images of parts
5 of the scene of Figure 1;

Figure 4 shows a schematic perspective view of a system for display of the moving images captured from the scene of Figure 1, following broadcast and then
10 receipt;

Figure 5 shows a more detailed block diagram of the second part of the system shown in Figure 3;

Figure 6 shows a schematic diagram of a second typical wide angle scene to be captured, broadcast and
15 displayed;

Figure 7 shows a block diagram of a first part of the system according to a second embodiment of the invention, for capture and transmission of images of the parts of the scene of Figure 6; and

Figure 8 shows a block diagram of a second part of the system according to the second embodiment of the invention for reception and display of images of parts of the scene of Figure 6.
20

Figure 1 shows, in schematic form, a typical wide angle scene 5 to be captured, broadcast and displayed.
25 The scene 5 is, in the example, in landscape format and comprises a background region 10 which is largely static over time and a plurality of smaller moving objects 20, such as human beings or vehicles, in the foreground.
30

The scene is captured as an image with a plurality of High Definition Television (HDTV) movie cameras 30; in the present example three are used. In

Europe, the HDTV Standard is defined in "ISO/IEC 13818-2; 1996 Information Technology - Generic coding of moving pictures and associated audio information - Part 2 : Video" The three cameras 30', 30" and 30''' each have fixed focal length lenses and are locked together. Moreover, each camera is locked in position during filming. Thus, over a period of time, the background region 20 of the image captured by each camera remains constant, because the field of view as seen through each camera lens remains static.

Only the foreground movement of the moving objects 20 causes one frame captured by one of the cameras 30 to differ from subsequent frames captured by the same camera. The principle may be likened to traditional animation, in which foreground figures are animated against a fixed background by changing the position of the foreground figures relative to the background during successive frames.

Each camera 30 is aimed at a different segment of the scene 5. Usually, each camera is identical in terms of focal length and distance from the scene so that each camera captures about one third of the total field of view afforded by the three cameras 30. The first camera 30' captures the left part of the scene 5 (indicated generally at 40 in Figure 1). The second camera 30" captures the centre part of the scene 5 (indicated generally at 50) and the third camera 30''' captures the right part 60 thereof. As seen in Figure 1, there is a degree of spatial overlap 70 between the captured images obtained by the three cameras 30. This is to prevent lines forming at the joins between the three captured images when they are displayed adjacent one another after broadcast, as will be

explained below.

Figure 2 shows a block diagram of a first part of the system for capturing and transmitting images from the scene 5 shown in Figure 1. As explained above, the scene is captured in three spatially adjacent segments 40, 50, 60 using three fixed focus, fixed direction cameras 30. In addition, other sensory information 80 such as sounds may be recorded at the same time.

The moving images captured by the three cameras from the three segments of the scene are sent to three respective digital encoders. Filmed images from the left camera 30' in Figure 1 are sent to a first digital encoder 90', filmed images from the centre camera 30" are sent to a second digital encoder 90", and filmed images from the right-hand camera 30"' are sent to a third digital encoder 90"'. Each of these encoders converts the filmed image sent to it into digital data. After digital encoding, each encoder then compresses the digital data using a suitable compression algorithm. In the present example, MPEG-2 is preferred. As will be understood by those skilled in the art, MPEG allows compression of a data stream obtained from moving images by removal of duplicate information in successive frames of the moving image. Thus, if the total information content of successive images in a moving image is broadly similar, then a high degree of compression may be achieved with MPEG, since, broadly, only the small amount of movement between frames needs to be encoded as motion vectors.

As explained in relation to Figure 1, by using fixed focal length, fixed view cameras, a static background is filmed with the only movement between

frames being due to the smaller moving objects 20, in the foreground. Because of the particular type of scene which is captured, and the way in which it is captured, a very high degree of compression is possible.

After each separate moving image has been encoded and compressed by the encoders 90, the three resultant video signals are multiplexed together using a multiplexer 100. The other sensory information 80 is also multiplexed together with the three encoded, compressed moving image signals. The multiplexed audiovisual signal which is the output of the multiplexer 100 is fed to a modulator and frequency converter which prepares the signal for broadcast in a manner which will be well known to those skilled in the art. After that, the modulated, frequency converted signal is broadcast via any suitable medium such as terrestrial, cable, satellite or MMDS. Because of the high degree of compression possible with the system, the multiplexed audiovisual signal which represents three high definition moving images as well as sounds and other sensory information can all be broadcast using such techniques without requiring extra bit rate capacities. For example, the typical time-averaged size of an HDTV signal from a single camera, filming a normal scene including pans, zooms and so forth is between 18 to 20 Mbit/second. The maximum bit rate capacity for digital cable or satellite transmission in Europe is 27.5 Mbit/second. The maximum digital terrestrial bit rate is slightly lower at 24 Mbit/second.

With the techniques described above, three HDTV signals can be encoded and compressed together with

audio information, and the total time-averaged size is still typically only around 12 Mbit/second, half or less than half of the maximum bit rate capacity currently available.

5 Figure 3 shows the second part of the system for reception and display of the broadcast audiovisual signal shown towards the right of Figure 2. In Figure 3, the broadcast signal is received and demodulated at a receiver/demodulator 120 and is then passed to a
10 demultiplexer 130. Here, the multiplexed signals are split up once more into the three video signals, and any other sensory information. The three video signals are sent to three separate high definition digital decoders 140', 140'', and 140'''. If the multiplexed
15 signal also contains audio information, then this is sent to an audio decoder 150. As set out above, yet further sensory information 160 may be included in the broadcast signal as well, and this is handled appropriately as will be described in connection with
20 Figure 4.

 The three high definition digital decoders 140 each decompress the data they receive to generate a digital data stream representative of the left, centre and right segments of the original filmed scene
25 respectively. Each of these video data streams is passed as an input to a processor 170 which decodes the three signals and combines them for displaying next to each other. The processor 170 also ensures that the edge of each image overlaps correctly with
30 the edge of an adjacent image such that the displayed composite image does not appear to have seams or joins in it.

 There are several factors to be considered in the

image processing to allow the display of a realistic image without apparent joins. Firstly, the screen shape needs to be considered. As described below, if a multiprojector image and a wide screen is employed, then it is preferable that the screen should be curved. More specifically, if the bulk of any foreground movement is captured by the centre camera 30", then that part of the screen onto which the resultant images captured by that camera are projected is typically flat. The side parts of the original scene, captured by the left and right cameras 30', 30'' respectively, will typically have little or no foreground or background movement and may be projected onto curved side parts of the screen.

To display the whole composite image correctly under these circumstances, it is usually necessary to process each image to correct for any screen curvature, and also to take into account the relative locations of the three cameras, except when the images are projected onto a right cylindrical screen. Then, a geometrically correct composite image may be obtained by locking the three cameras at predetermined angles relative to one another.

The techniques for processing images when the screen is not a pure cylinder do not form part of the present invention and will not be described in further detail. However, it is to be noted that the composite signal which is broadcast may further include information relating to the lens aberrations of each camera. This information is employed by the image processor to help dewarp the projected image.

Once the correct image geometry has been obtained to generate a realistic composite image, edge

alignment processing is carried out. This is accomplished by block matching pixels according to their spatial position, for example by searching for Moiré patterns in the frequency domain by carrying out a Fourier transform. Finally, contrast match and colour match processing is carried out to avoid shadows or other artefacts from being generated.

One technique is to try and match vertical bands, for example through linear interpolation of the luminance in the various bands. To avoid excessive luminance in the areas of overlap of spatially adjacent images, the projectors (see below) may be controlled so as to project at half luminance over the area of overlap (so that the total luminance in that area is normal).

The particular preferred combination of fixed view, fixed focus cameras and MPEG-2 encoding also allows heavy noise suppression to be successfully applied. The basic principle of video noise suppression is that there is a certain amount of correlation between the video content of successive frames, whereas the noise content is essentially random from frame-to-frame with no correlation.

Because the video content of successive frames captured with fixed view, fixed focus cameras is substantially identical (with only a small amount of foreground movement, as previously described), a noise suppression filter may be employed which utilizes this feature to suppress noise. In general terms, the amount of noise in the eventually displayed images is inversely related to the amount of "averaging" between successive images which is possible. The amount of averaging, in turn, is related to the amount of

movement between successive frames. Thus, the small amount of difference between successive images in the present case allows significant noise reduction.

5 There are a number of ways of implementing noise reduction. One preferred technique employs recursive feedback. A frame store latches an image at time t and acts as a delay, with the output being fed back to the input via an attenuator, most typically a multiplier for digital signals. For successive images
10 at times $t + \Delta t$, having significant movement between them, the amount of recursion must be reduced, else trails or smears appear on the decoded film. However, when the successive images are very similar, recursion will be large, that is, the eventually displayed film
15 is the average of many successive frames, which heavily suppresses noise.

Two different display mechanisms are supported by the system of Figures 2 and 3. In a first mode, the combined composite image generated by the processor
20 170 is displayed using a multiprojector arrangement. This is shown schematically in Figure 4. The output of the processor 170 in Figure 3 controls three HDTV projectors 180, 190, 200. In the illustrated embodiment of Figure 4, each of the projectors is an
25 LCD or cathode ray tube (CRT) projector which back-projects onto a large screen 210. As previously explained, the left-hand projector 180 projects images captured from the left-hand third 40 of the scene shown in Figure 1. Likewise, the centre projector 190
30 projects images captured from the centre third 50 of the scene 5 of Figure 1, and the right-hand projector 200 projects images from the right-hand third 60 of the original scene. Although the screen 210 is shown

with vertical dotted lines, this is merely for the purposes of explanation and it will be understood that, in practice, the processor operates so as to ensure a seamless blend between the left, centre and right images on the screen as explained previously. The curved nature of the screen 210 may also be seen in Figure 4.

The display of Figure 4 may also include loudspeakers 220. Although only two are shown in Figure 4, a plurality of loudspeakers may be used to play back multitrack audio recordings to enhance the sense of realism. The loudspeakers 220 are fed with an audio signal 230 obtained from the audio decoder 150 of Figure 3.

As previously mentioned yet other sensory information can be provided further to increase the sense of realism. For example, a fan 240 may be supplied with trigger signals sent with the broadcast audiovisual signal to generate wind at appropriate moments depending upon the wind at the original scene 5. A viewer 250 is typically seated in front of the screen 210, facing the fan 240, and the seat 260 upon which the viewer 250 sits may itself be mounted upon a movable platform 270. This movable platform 270 may respond to further sensory information 160 derived from the original scene 5 to represent ground movement, for example.

Of course, the number of projectors does not have to coincide with the number of cameras employed to capture the original image. Indeed, a single projector may project all of the images.

As an alternative to the multiprojector system of Figure 4, a head-mounted display may instead be

employed. The general principles of head-mounted displays will be familiar to those skilled in the art. In this case, a helmet device is placed upon the viewer's head such that a screen wraps around the viewer's eyes. Headphones are mounted into the helmet around the ears. Again, the composite processed image from the processor 170 of Figure 3 is fed to the head-mounted display for display upon the screen thereof. In this case, however, only a portion of the composite image is output at any one time to the display of the head-mounted display. It will be understood, however, that the whole composite image is still captured broadcast and received for processing. However, it is the position of the user's head relative to his body which defines which part of the composite image is displayed on the screen of the head-mounted display. Thus, by moving his head from left to right, the user is given the impression that he is looking around the original film scene 5. If a head-mounted display is used, it is also preferable that the system includes audio processing 175 to provide phased audio so that, as the user moves his head, the source of sound relative to the viewed image appears to remain stationary.

Figure 5 shows the features of the block diagram of Figure 3 in more detail. Following demodulation of the demodulator 120, the broadcast signal is demultiplexed into three separate video signals using the demultiplexer 130 (the audio signals are not shown for the sake of clarity in Figure 5). The demultiplexer 130 is controlled by a system controller 280. The demultiplexer 130 also recovers the clock signal so that the three resultant images can be

synchronised which is important if the composite image which is ultimately displayed is to appear realistic. Clock recovery is accomplished, in the embodiment of Figure 5, using separate hardware indicated at 290.

5 Following demultiplexing, the three signals are separately decompressed at the decoders 140', 140" and 140'''. The outputs of the three decoders are fed to separate respective HDTV field memory devices 151, on a separate printed circuit board (pcb), for storage.

10 The separate pcb also contains a sync processor 152 which allows time synchronization of frames captured at the same time by the three cameras.

 The remainder of the separate pcb contains hardware for carrying out the image processing as previously described. In the embodiment of Figure 5, 15 separate overlap and brightness processors 153, 154 adjust the brightness and control the overlap regions of the centre and left images, and centre and right images respectively.

20 The outputs of the two overlap and brightness processors 153, 154 are sent to the three projectors 180, 190 and 200 of Figure 4.

 If a head mounted display is to be employed, current technology only allows a 'standard' television 25 image to be displayed. Thus the HDTV images are converted into standard format and stored in an SDTV field memory 156 before display. An upper and lower border memory 157 is also employed in that case.

 The second part of the system shown in Figure 3 30 and, in more detail, in Figure 5 can suitably be provided as a single "set top box" with all of the demodulation, decoding, decompressing and processing being carried out using solid state electronics. The

set top box may also receive further information from the cameras which capture the original scene 5 relating to camera lens aberration, to assist in edge processing as described above.

5 Figure 6 shows a second scene 300 which may be recorded using a single camera and which allows a large, wide screen image to be captured, broadcast and displayed using a relatively low bit rate capacity. The scene 300 shown in Figure 6 is of a football
10 match. Once again, a fixed focal length, fixed view camera is employed, but the scene 300 is notionally segmented into three separate portions 310, 320 and 330. The first portion 310 contains the sky above and behind the stadium in which the televised game is
15 played. The second portion 320 contains the stands with the crowd in them. The third portion 330 contains essentially the pitch and its immediate surroundings.

A sense of realism can be generated by broadcasting only the third portion 330, containing
20 the football pitch and players, in real time, and displaying animated (non-real-time) images of the crowd and sky at the same time. The advantage of this is that the total bit rate capacity required to broadcast such an image is dramatically reduced.

25 Figure 7 shows a block diagram of the first part of the system for capturing and transmitting images from the scene 300 of Figure 6. A high definition camera 400 first captures a short (typically a few
30 seconds) length of film from the first portion 310 of the image 300, including the sky. This short length of film is encoded using an encoder 410, is then optionally compressed, and then passed to a modulator/frequency converter to allow broadcast.

After broadcast of the short length of film captured from the first portion 310 of the image 300, the camera is then moved to capture a short length of film from the second portion 320 of the image 300, containing the stadium and the crowd. Once again, the short length of captured images is encoded with the digital encoder 410, optionally compressed, and then modulated and frequency converted using the modulator, frequency converter 420 before being broadcast.

Once the first and second portion 310, 320 of the image 300 has been captured and broadcast, the camera 400 is aimed at, and locked in position, such that it captures only the third portion 330 of the image 300, containing the football pitch, players and so forth. Once the game commences, the camera 400 generates a substantially continuous "live feed" of moving images which are encoded by the digital encoder 410, modulated and frequency converted with the modulator/frequency converter 420 and then broadcast. Continuous capture of images occurs until the game finishes, for example. As with the scene 5 of Figure 1, the use of a fixed focal length, fixed view camera for capturing the live images from the third portion 330 of the image 300 allows a high degree of image compression (for example using MPEG) to be achieved. This is because the pitch, goals, flags and so forth are largely stationary from one image to the next, and it is only the players and ball which move around the pitch and which thus contribute to a change between successive images.

Although, in Figure 7, only one high definition camera 400 is employed, with the three separate portions 310, 320, 330 of the image 300 being captured

sequentially, it would of course be possible to use separate cameras 400' and 400" to capture images from the three portions, rather than using one camera which must be moved to capture images from the three
5 separate portions.

Figure 8 shows a block diagram of the second part of a system for capture, broadcast and display of the scene shown in Figure 6. Once the first portion 310 of the scene 300 has been captured and broadcast, it
10 is received by a receiver/demodulator 430. After that, it is decoded by a decoder 440. The short film is then sent along a first line 450 to a memory device 460 which stores the array of images making up the short film for future use. The memory device 460 can
15 either store the images in compressed form or can first decompress them.

Once decoding and storing of the images obtained from the first portion 310 of the image 300 have been completed, images captured from the second portion 320
20 of the scene 300 and then broadcast are next received and demodulated at the receiver/demodulator 430. Again, the signal is decoded by the decoder 440 and then sent along line 450 to the memory device 460 for storing.

25 Finally, once this stage has been completed, the system is ready to receive the continuous stream of images captured from the third portion 330 of the image 300 and broadcast as explained in connection with Figure 7 above. This time, however, after
30 reception and demodulation at the receiver/demodulator 430 and decoding by the decoder 440, the signal is passed along a second line 470 to an image processor 480. Once the image processor starts to obtain the

continuous "live" images from the third portion 330 of the scene 300, both the short film captured from the first portion 310 and the short film captured from the second portion 320, which are stored in the memory device 460, are sent to an animation processor 490. The animation processor receives each of the two separate films and animates them in an endless loop. The endless loop of images from the first and second portions 310, 320 is then passed to the image processor 480 which combine all three images (the animated films from the first and second portions 310 of the scene 300, and the continuous "live" images from the third portion 330) to produce a composite image. This composite image may, as with the system of Figures 1 to 5, use overlap processing to ensure that there is no appearance of a join between the three moving images.

At this point, the display of the composite image may be achieved using the same arrangement as is shown in Figure 4. In particular, the image processor 480 can control the three separate projectors 180, 190, 200 to project the three separate moving images from the three portions 310, 320, 330 of the scene 300 respectively. Of course, rather than being arranged horizontally as shown in Figure 4, with the scene of Figure 6 it is preferable that the three projectors 180, 190, 200 are instead arranged vertically, with the upper projector projecting the animated images from the first portion of the scene 300 (the sky), the middle projector projecting images of the second portion 320 of the scene 300 (the crowd and stadium), and the lower projector projecting images obtained from the third portion 330 of the scene 300 (the

"live" images of the pitch and players). However, it will be understood that the choice of vertical or horizontal segmentation of the captured scene will be dependent upon the nature of that scene and a scene with foreground movement occurring along a generally vertical band of that original scene might preferably be segmented horizontally instead.

Once again, a separate live audio signal can be sent together with the live video images obtained from the third portion 330 of the scene 300. Likewise, further sensory information such as wind may be sent as well.

The system described in connection with Figures 6, 7 and 8 is of course subject to various modifications. Rather than pre-broadcasting only one short film from the first and second portions 310, 320, a number of short films can be pre-broadcast instead and each of these can be stored separately in the memory device 460. Then, an appropriate one of the stored short films in the memory 460 can be triggered and animated as appropriate. For example, if a goal is scored, the image processor may obtain a short film of an appropriate crowd reaction from the memory device 460 for animation.

Likewise, although pre-broadcast of the short films to be animated, from the first and second portions 310, 320 of the image 300 respectively, permits a signal well within the bit rate capacity of current broadcasting systems to be achieved, it is of course possible to obtain such short images for concurrent broadcast whilst the live video images from the third portion 330 of the image 300 are being broadcast. In this case, of course, three separate

cameras 400, 400' and 400" would be necessary.
Because only short amounts of film would be sent,
lasting no more than a few seconds, the bit rate
capacity of current broadcasting systems would still
5 be sufficient.

Finally, it is to be appreciated that the two
systems described in connection with Figures 1 to 5
and 6 to 8 respectively are not mutually exclusive. In
particular, an ultra-wide image could still be
10 successfully captured and broadcast using current
broadcasting systems by, for example, segmenting a
scene to be captured and broadcast into nine segments
(three rows x three columns). The first row could
comprise three spatially adjacent animated images of
15 three spatially adjacent parts of the sky. The second
row of three spatially adjacent images could contain
three spatially adjacent parts of the crowd, and the
bottom row could contain three spatially adjacent
images, each compressed as explained in connection
20 with Figure 1, and then combined at the
receiver/display. Likewise, the noise suppression
techniques described in connection with the first
embodiment are equally applicable to the second
embodiment.

25 The techniques described above are, of course,
not restricted to live events such as football matches
or concerts. Using the well known "chromakey"
technique, moving objections (usually actors) can be
overlaid onto a background. By using a stationary
30 background, as explained above, the overall
compression of a chromakey image where only the actor
moves can be substantial.

Finally, whilst MPEG-2 has been described as the

preferred compression algorithm, MPEG-4 may be employed instead. MPEG-4 is particularly suitable for segmented images.

CLAIMS

1. A method of capturing, broadcasting and displaying a moving image, comprising:

- 5 (a) capturing, with n separate cameras, where n is a positive integer greater than 1, n respective separate arrays of temporally successive still image subframes, each of the n arrays being captured from separate spatially adjacent subdivisions of the moving
10 image respectively;
- (b) digitally encoding said n captured arrays to generate a corresponding number, n , of separate encoded arrays, each of which contains subframe image
data;
- 15 (c) compressing the size of at least some of the n digitally encoded arrays;
- (d) multiplexing the n encoded arrays to form a composite encoded image signal;
- (e) transmitting the composite encoded image
20 signal;
- (f) receiving the composite encoded image signal;
- (g) demultiplexing the composite encoded image signal to reconstitute the said number, n , of encoded
25 arrays;
- (h) decoding and decompressing the n encoded, compressed arrays to reconstitute the n captured arrays obtained from the original moving image; and
- 30 (i) displaying the n captured arrays in spatial order such that a facsimile of at least a spatial part of the original moving image may thereby be represented; the subframe image data of the encoded arrays being such that, over a given time period, the

sum of the compressed sizes of the arrays is the same as or less than the size of the data contained within the largest of the n uncompressed arrays over that time period.

5

2. The method of claim 1, in which each of the n arrays is compressed using MPEG-2 compression.

10 3. The method of claim 1 or 2, in which in a given one of the n arrays, the subframe image data of successive ones of the encoded temporally successive image subframes, within that array and prior to compression, is substantially the same.

15 4. The method of claim 1, claim 2 or claim 3, further comprising averaging the subframe image data over temporally successive image subframes to reduce the level of random noise encoded therein.

20 5. The method of any preceding claim, comprising providing 3 separate cameras, each arranged to capture, spatially, substantially one third of the moving image.

25 6. The method of any preceding claim, further comprising:
capturing an audio soundtrack in synchronism with the n arrays of temporally successive image subframes;
digitally encoding the said audio soundtrack;
30 multiplexing the encoded audio soundtrack with the n encoded arrays such that the resultant composite encoded image signal includes both audio and video data;

demultiplexing the composite encoded image signal after the said steps (e) and (f) of transmitting and receiving, to reconstitute both the n encoded arrays and the encoded audio soundtrack;

5 decoding the said audio soundtrack; and playing back the said decoded audio soundtrack in synchronism with the n arrays of moving images.

10 7. A system for capturing, broadcasting and displaying a moving image, comprising:

 a plurality, n , of separate cameras, where n is a positive integer greater than 1, arranged respectively to capture n separate arrays of temporally successive still image subframes, each of the arrays being
15 captured from separate spatially adjacent subdivisions of the moving image respectively;

 an encoder arranged to digitally encode the n captured arrays, to generate a corresponding number, n , of separate encoded arrays, each of which contains
20 subframe image data;

 a processor, arranged to compress the size of at least some of the n digitally encoded arrays;

 a multiplexer, arranged to multiplex the n encoded arrays to form a composite encoded image
25 signal;

 a transmitter arranged to transmit the composite encoded image signal;

 a receiver arranged to receive the composite encoded image signal;

30 a demultiplexer for demultiplexing the composite encoded signal to reconstitute the said number, n , of encoded arrays;

 a decoder arranged to decode and decompress the n

encoded, compressed arrays to reconstitute the n
captured arrays obtained from the original moving
image; and

5 a display for displaying the n captured arrays in
spatial order such that a facsimile of at least a
spatial part of the original moving image may thereby
be represented; the subframe image data of the encoded
arrays being such that, over a given time period, the
sum of the compressed sizes of the arrays is the same
10 as, or less than, the size of the data contained
within the largest of the n uncompressed encoded
arrays over that time period.

8. The system of claim 7, in which each of the
15 n separate cameras is a movie camera arranged to
capture the respective separate array of moving
images.

9. The system of claim 7 or claim 8, in which
20 the processor is adapted to compress each of the n
arrays using MPEG-II.

10. The system of any one of claims 7 to 9, in
which, in a given one of the n arrays, the subframe
25 image data of successive ones of the encoded
temporally successive image subframes, within that
array and prior to compression, is substantially the
same.

30 11. The system of claim 7, claim 8 or claim 9,
further comprising a noise filter arranged to generate
an average of the subframe image data over temporally
successive image subframes to reduce the level of

random noise encoded therein.

12. The system of any of claims 7 to 11, in which each of the n cameras has a lens of fixed focal length, and in which each camera is fixedly mounted such that, in each of the n arrays, the subframe image data of a given one of the subframe images is substantially the same as the subframe data of a temporally successive subframe image within a given one of the n arrays.

13. The system of any one of claims 7 to 12, comprising 3 separate cameras each arranged to capture, spatially, substantially one third of the image.

14. The system of any of claims 7 to 13, in which the n cameras are high definition (HDTV) cameras.

15. The system of any of claims 7 to 14, further comprising:

a microphone for capturing an audio soundtrack in synchronism with the n arrays;

a sound encoder for digitally encoding the said audio soundtrack; and

a sound reproduction device,

wherein the multiplexer is further arranged to multiplex the encoded audio soundtrack with the n encoded, compressed arrays such that the resultant composite encoded image signal includes both audio and video data, the demultiplexer is further arranged to demultiplex the composite encoded image signal

following reception thereof by the receiver, to reconstitute both the n encoded arrays and the encoded audio soundtrack, and the decoder is further arranged to decode the said audio soundtrack such that the sound reproduction means may play back the said decoded audio soundtrack in synchronism with the n arrays.

16. The system of any of claims 7 to 15, in which the display includes a plurality, n, of projectors and a screen, each of the n projectors being arranged to project a respective one of the n arrays onto a part of the screen, the n arrays being projected in the spatial order in which they were captured, such that the facsimile of substantially all of the original moving image is represented upon the screen.

17. The system of claim 16 when dependent upon claim 15, in which the n projectors are arranged to project an array in HDTV format onto the screen.

18. The system of any of claims 7 to 15, in which the display includes a head mounted viewing device arranged to display a spatially limited portion of the original moving image, movement of the head-mounted display device causing different spatially limited parts of the original moving image to be displayed.

19. A method of capturing, broadcasting, and displaying a moving image, comprising:

capturing, with a camera, a substantially

temporally continuous stream of first still image
subframes from a first spatial part of the moving
image, the stream of first still image subframes
representing a substantially temporally continuous
5 motion picture image;

digitally encoding the captured first still image
subframes;

capturing, with a camera, a finite series of
second still image subframes from a second spatial
10 part of the moving image, the finite period of time
being short in relation to the time over which the
substantially continuous stream of second still image
subframes is captured, the finite series of second
still image subframes representing a finite length
15 motion picture image;

digitally encoding the said finite series of
second still image subframes;

transmitting the said first and second encoded
still image subframes;

20 receiving the said first and second encoded still
image subframes;

decoding the first and second encoded still image
subframes;

providing a display;

25 displaying the said first still image subframes
within a first spatial part of the display such that
the substantially temporally continuous motion picture
image is represented; and

repeatedly displaying the said second still image
30 subframes within a second spatial part of the display
such that the finite length motion picture image is
represented as an animated loop;

the display thereby displaying a facsimile of at

least a spatial part of the original moving image.

20. The method of claim 19, in which the
digitally encoded finite series of second still image
subframes is transmitted and received prior to the
capture of the stream of first still image subframes,
the method further comprising storing the received
second still image subframes after transmission and
reception for subsequent display as an animated loop.

21. The method of claim 20, in which the same
camera captures the finite series of second still
image subframes and, subsequently, the stream of first
still image subframes.

22. A system for capturing, broadcasting and
displaying a moving image, comprising:

a first camera, arranged to capture a
substantially temporally continuous stream of first
still image subframes from a first spatial part of the
moving image, the stream of first still image
subframes representing a substantially temporally
continuous motion picture image;

a second camera, arranged to capture a finite
series of second still image subframes from a second
spatial part of the moving image, the finite period of
time being short in relation to the time over which
the substantially continuous stream of second still
image subframes is captured, the finite series of
second still image subframes representing a finite
length motion picture image;

an encoder arranged to digitally encode the
captured stream of first still image subframes and to

digitally encode the captured finite series of second still image subframes;

a transmitter for transmitting the said first and second encoded still image subframes;

5 a receiver for receiving the said first and second encoded still image subframes;

a decoder arranged to decode the first and second encoded still image subframes;

a display having first and second spatial parts,
10 the system being arranged to display the said first still image subframes within the first spatial part of the display such that the substantially temporally continuous motion picture image is represented and to repeatedly display the said second still image
15 subframes within the second spatial part of the display such that the finite length motion picture image is represented as an animated loop; the display thereby displaying a facsimile of at least a spatial part of the original moving image.

20

23. The system of claim 22, in which the encoder receives and encodes the finite series of second still image subframes prior to receipt and encoding of the stream of first still image subframes, and in which
25 the transmitter transmits the second still image subframes and the receiver receives the second still image subframes before transmitting and receiving the first still image subframes respectively, the system further comprising:

30 a storage device local to the receiver, the storage device being arranged to receive and store the said finite series of second still image subframes; and

a processor arranged to receive the finite series of second still image subframes from the storage device and to animate the said finite series as an animated loop in synchronism with the temporally continuous motion picture subsequently received by the receiver.

24. A method of capturing, broadcasting and displaying a moving substantially as herein described with reference to the Figures.

25. A system for capturing, broadcasting and displaying a moving image substantially as herein described with reference to the Figures.



Application No: GB 9916337.0
Claims searched: 1-25

Examiner: Iwan Thomas
Date of search: 10 January 2001

Patents Act 1977 Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.S): H4F FAAE, FAAX, FGG, FGH, FGS, FGT

Int Cl (Ed.7): H04N 5/262

Other: Online: WPI, EPODOC, JAPIO

Documents considered to be relevant:

Category	Identity of document and relevant passage	Rélevant to claims
A	GB 2050748 A (HONEYWELL) See whole document	
A	GB 2014015 A (HONEYWELL) See whole document	
A	EP 0276800 A2 (IBP) See col. 3 line 52 - col. 4 line 1	

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.